Data Analytics

Rick Archibald & Hoang Tran
Oak Ridge National Laboratory

Data Analytics for the DOE Large Scale Scientific Instruments

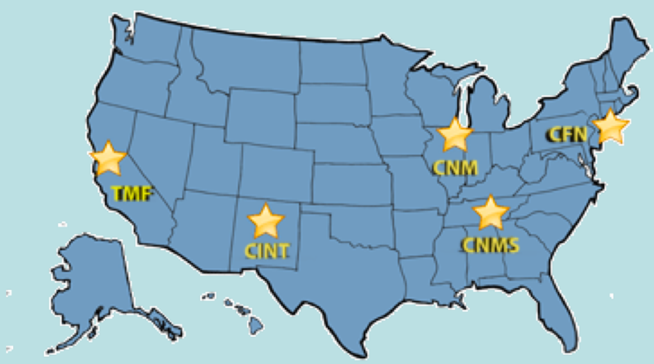
Experimental Facilities

SNS
SPALLATION NEUTRON SOURCE

ALS
ADVANCED LIGHT SOURCE

HFIR

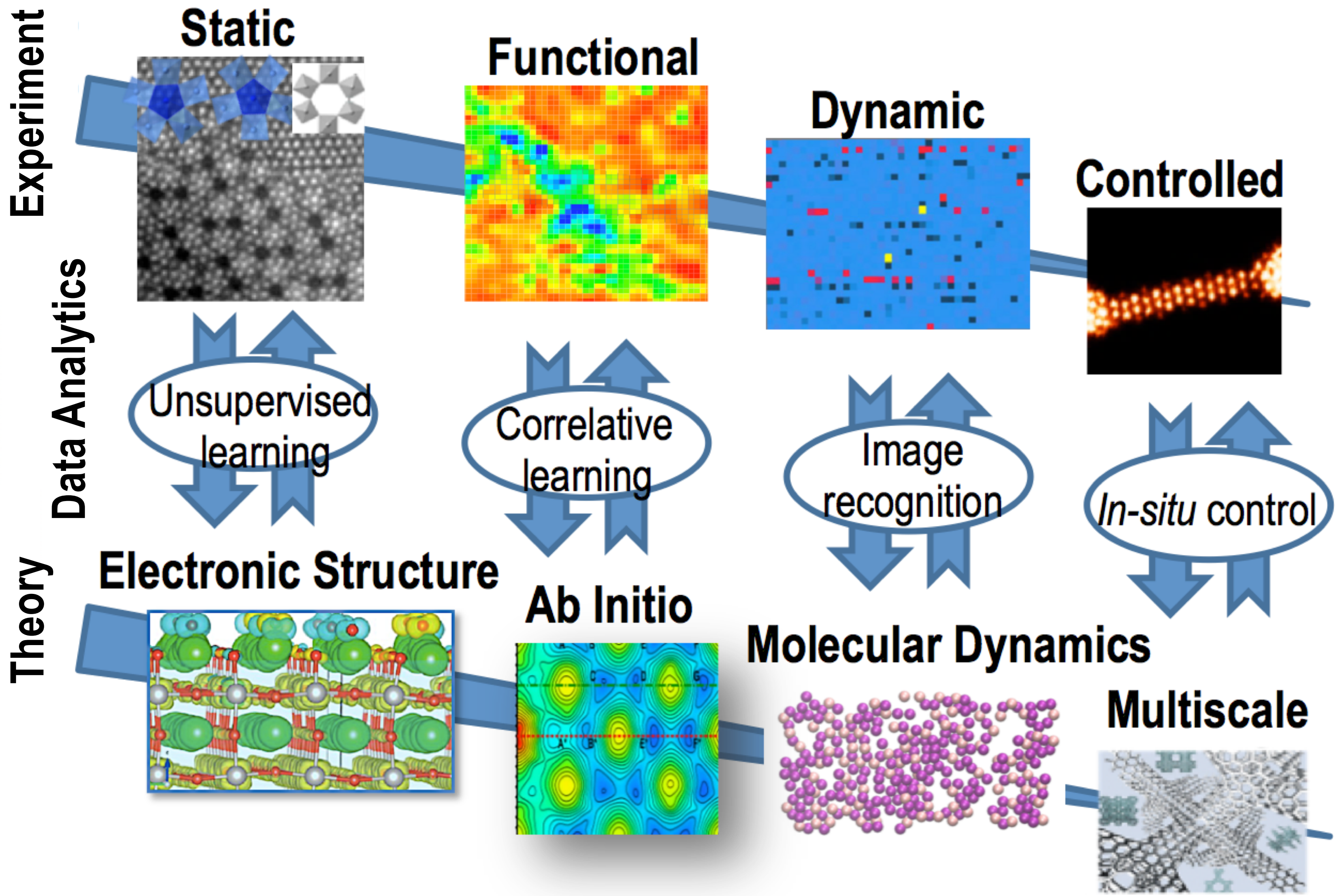
JGI
JOINT GENOME INSTITUTE
DEPARTMENT OF ENERGY



Computational Facilities

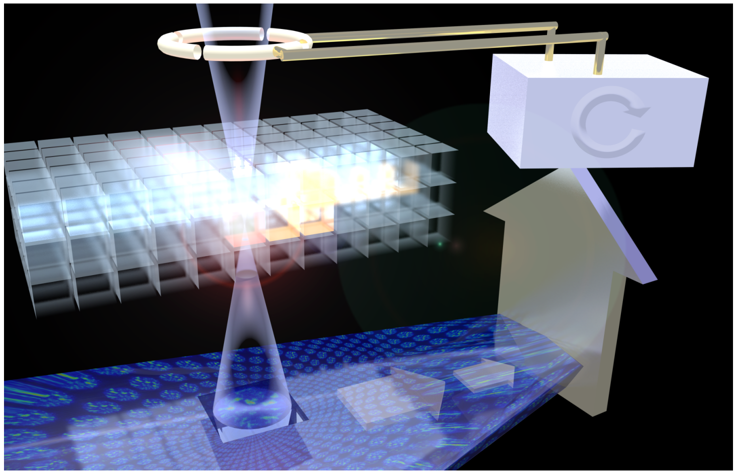
FASTMATH

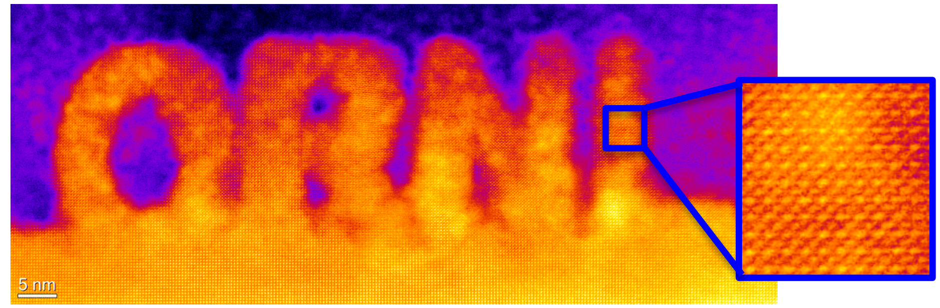
Data Analytics Bridge Experiment and Theory

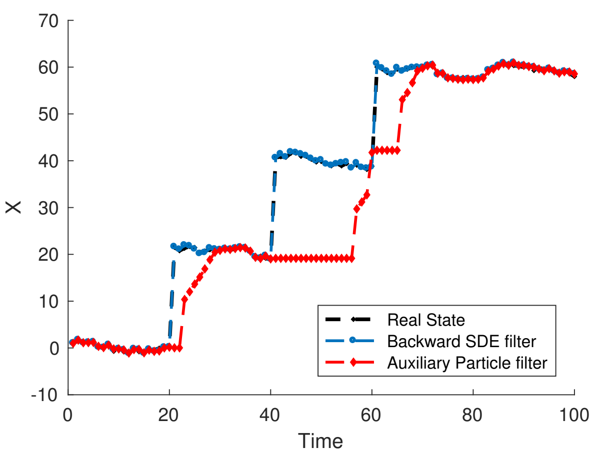


FASTMATH

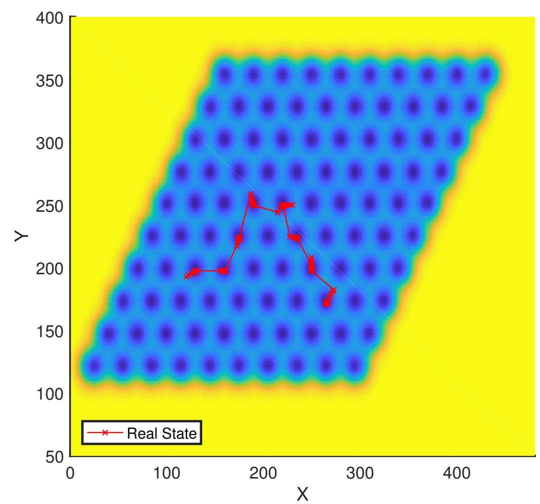
Atomic Forge







Tracking of multiple well potential in 1D.



Tracking of multiple well lattice potential in 2D

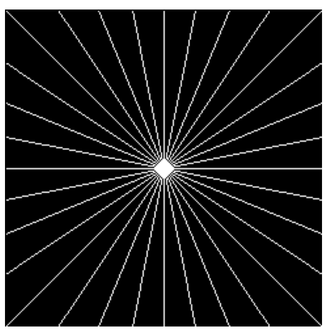
- Need fast accurate forward models for atomic states of tracked atoms.
- Need computational control of beam position and intensity.
- Will enable 3D atomic fabrication: quantum computing, spintronics, etc.

General Framework for Data Reconstruction (connection to FASTMath Optimization and UQ)

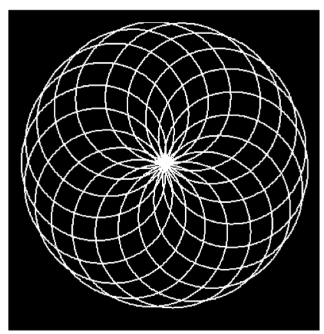
Determine $\mathbf{f} = \{f(x_i, y_j) : 0 \leq i, j \leq 2N\}$ that solves the convex optimization problem

$$\begin{aligned} \text{minimize} \quad & ||J_x \mathbf{f}||_1 + ||J_y \mathbf{f}||_1, \\ \text{subject to} \quad & ||MF\mathbf{f} - \hat{\mathbf{f}}||_2 \leq \sigma, \end{aligned}$$

Where the matrix M is a mask that removes unknown Fourier coefficients.



Tomography

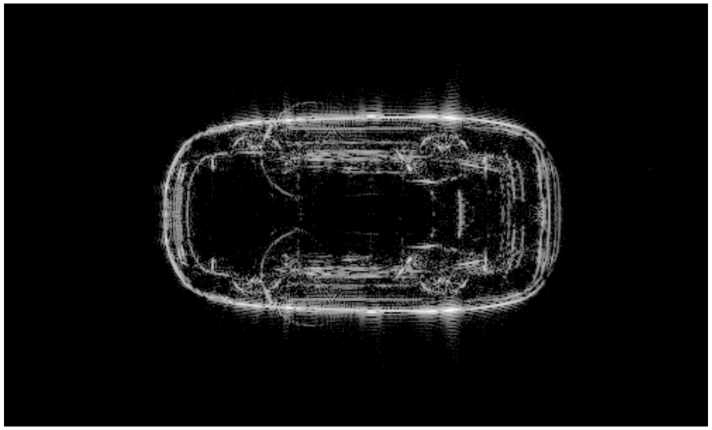


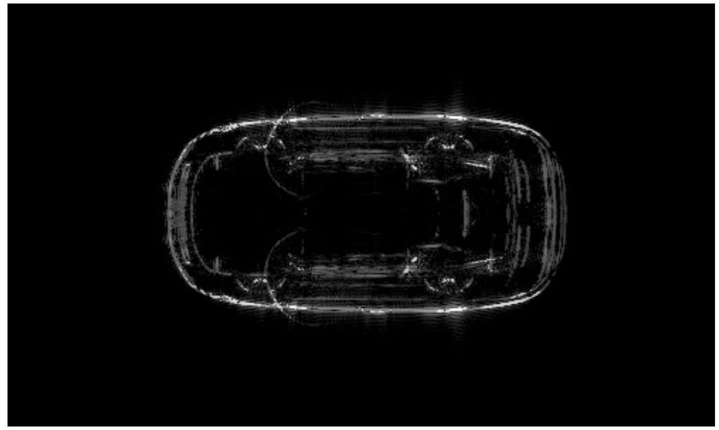
MRI

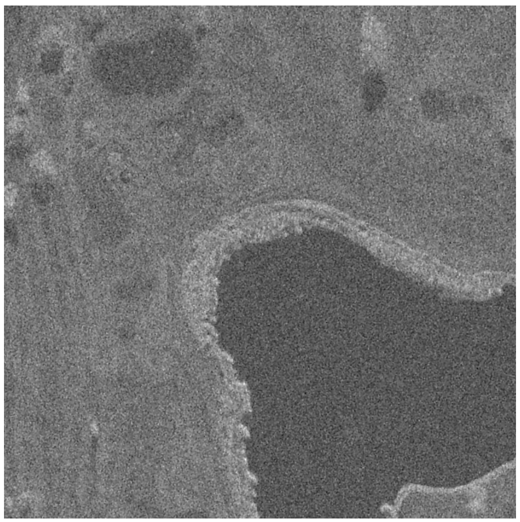
Ultrasound

FASTMATH

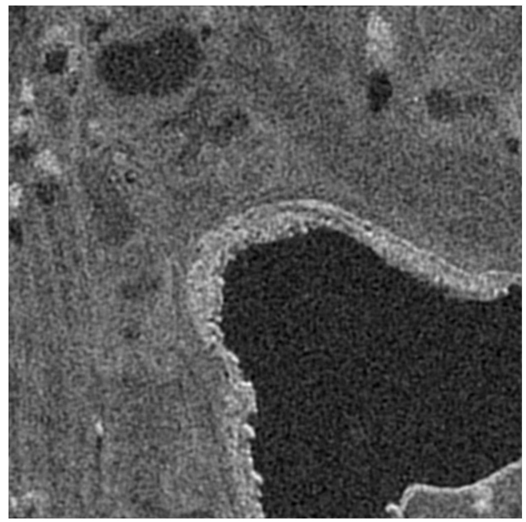
SAR Reconstruction Results







Improving synthetic aperture radar (SAR) data through AHOTV. Left TV SAR and right AHOTV reconstruction of car and Golf Course



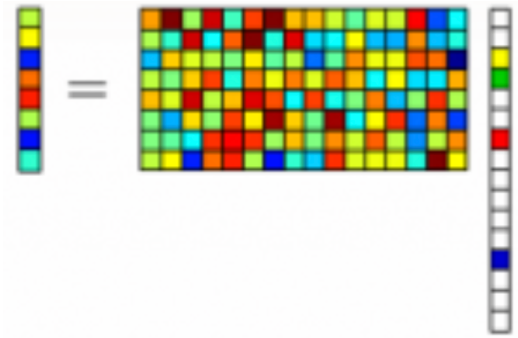
1. Brugiapaglia, Adcock, and Archibald, “Recovery guarantees for compressed sensing with unknown errors”, *Sampling Theory and Applications*, 2017.
2. Churchill, Archibald, and Gelb, “Edge-adaptive l2 regularization image reconstruction from non-uniform Fourier data”, *Journal of Scientific Computing*, 2018.

FASTMATH

Sparse reconstruction and representation of data

- We reconstruct data $c \in \mathbb{R}^{N\times l}$ from measurements $u \in \mathbb{R}^{m\times l}$ and $A \in \mathbb{R}^{m\times N}$:

$$u \approx Ac$$



  - Limited number of measurements: $m \ll N$.
  - The data are sparse.
  - $l = 1$: reconstructing a single dataset.
    $l > 1$: simultaneously reconstructing multiple datasets.

- Recovery via regularizations enforcing sparsity:

$$c = \text{argmin } R(z) \quad \text{subject to} \quad u \approx Az$$

Standard CS: $R(z) = \|z\|_1$.

Structures of the sparsity can be exploited:

- Downward closed and tree structures: $R(z) = \|z\|_{\omega,1}$.
- Joint sparsity: $R(z) = \|z\|_{2,1}$.

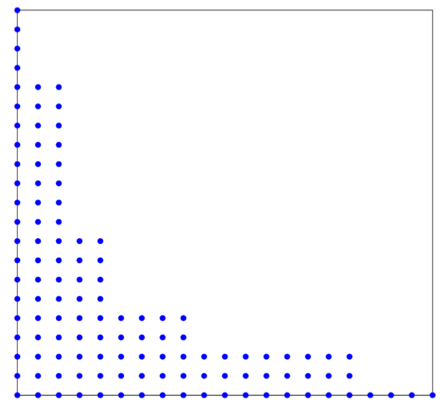
FASTMATH

Sparse reconstruction and representation of data

- Data from UQ and imaging applications often possess downward closed and tree structure.

- Weighted l_1 minimization with a specific choice of weight:

$$R(z) = \|z\|_{\omega,1} \text{ with } \omega_j = \max |A_{:,j}|$$



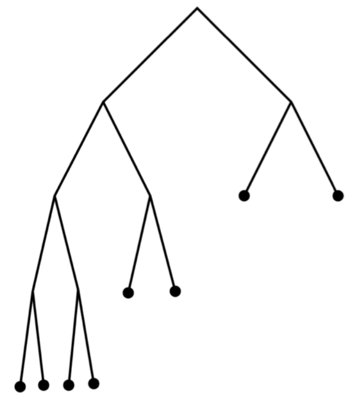
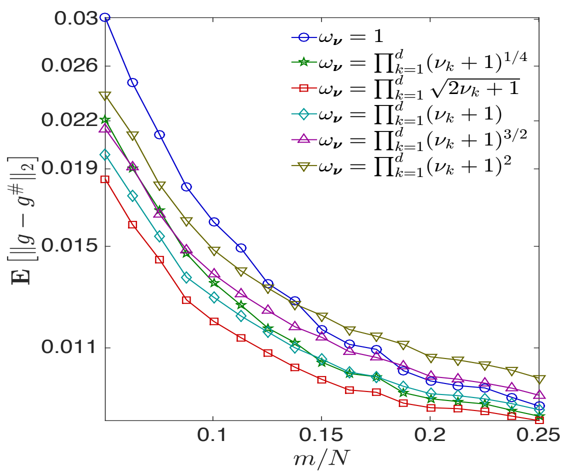


Figure: *A comparison of weighted l_1 minimization with different choices of weights*



Certified reduction in complexity:

- Legendre systems: $m = O(s^2)$ instead of $O(s^{2.58})$ as in unweighted l_1.
- Chebyshev systems: $m = O(s^{1.58})$ instead of $O(s^2)$ as in unweighted l_1.

A. Chkifa, N. Dexter, H. Tran, and C. Webster, *Polynomial approximation via compressed sensing of high-dimensional functions on lower sets*. **Math. Comp.** (2017) https://doi.org/10.1090/mcom/3272

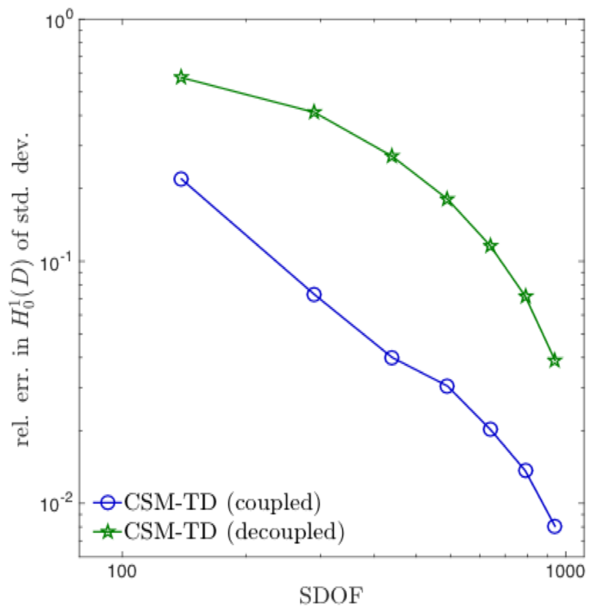
Sparse reconstruction and representation of data

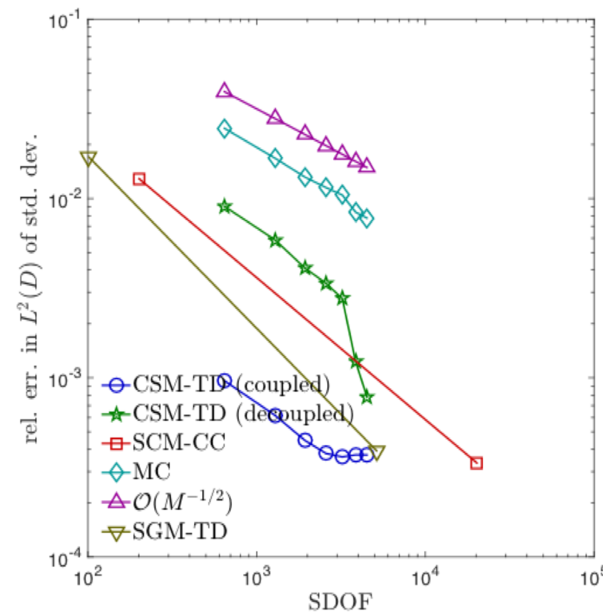
- Simultaneous reconstruction of multiple datasets sharing similar sparsity patterns using mixed norm:

$$R(z) = \|z\|_{2,1} = \sum_{j=1}^{N} \|z_{j,:}\|_2$$

  - $\|.\|_{2,1}$ promotes the joint sparsity of column vectors.
  - Provably yielding better recovery properties than individual reconstructions.
  - Efficiently implemented by proximal splitting approaches.

Figure: *A comparison of joint sparse with individual reconstructions as well as other techniques in approximating high dimensional parameterized systems.*





N. Dexter, H. Tran, and C. Webster, *On the strong convergence of forward-backward splitting in reconstructing jointly sparse signals.* **submitted,** 2017. https://arxiv.org/abs/1711.02591

FASTMATH

Compression Artifact Removal in Scientific Data Using Deep Learning (Connection to RAPIDS)

Scientific Achievement

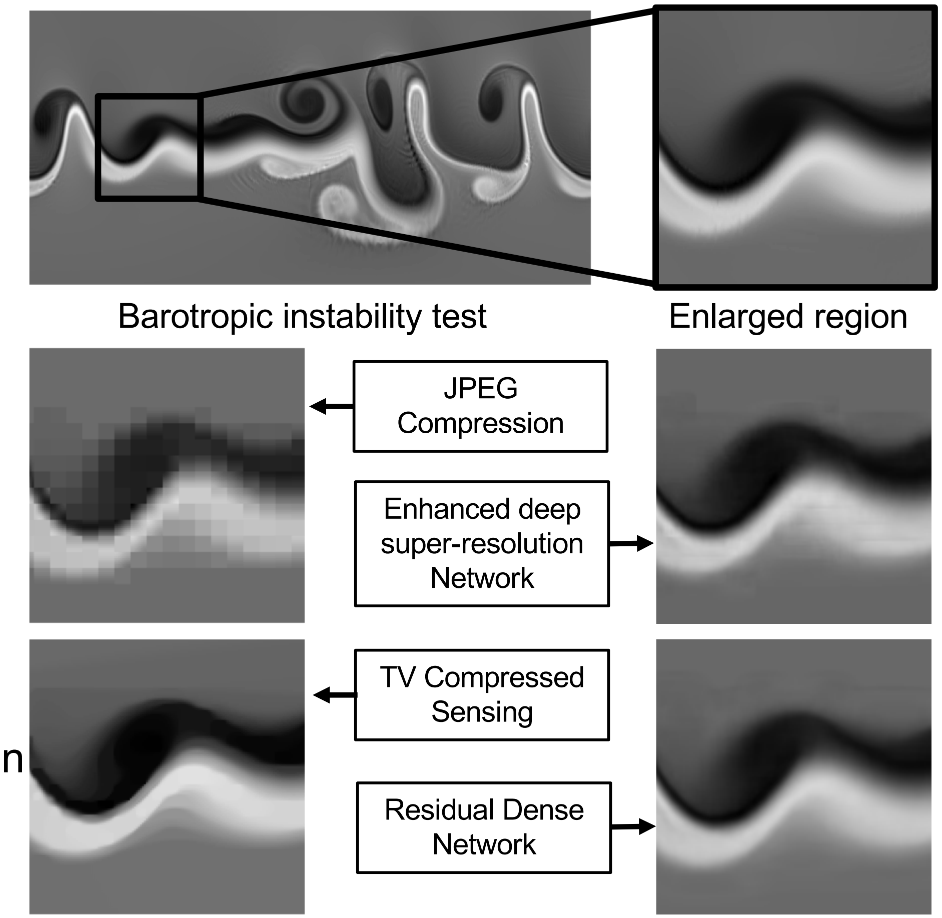
Developed a deep-learning based compression artifact removal approach that provides f**ast enhancement (using trained model) compared to state-of-the-art** compressed sensing (CS) approach

Significance and Impact

Scientific simulations generate large amounts of data. Storing/moving it can be expensive, and lossy compression like JPEG results in compression artifacts (ringing, blocking, etc.). CS is expensive and fails in some cases.

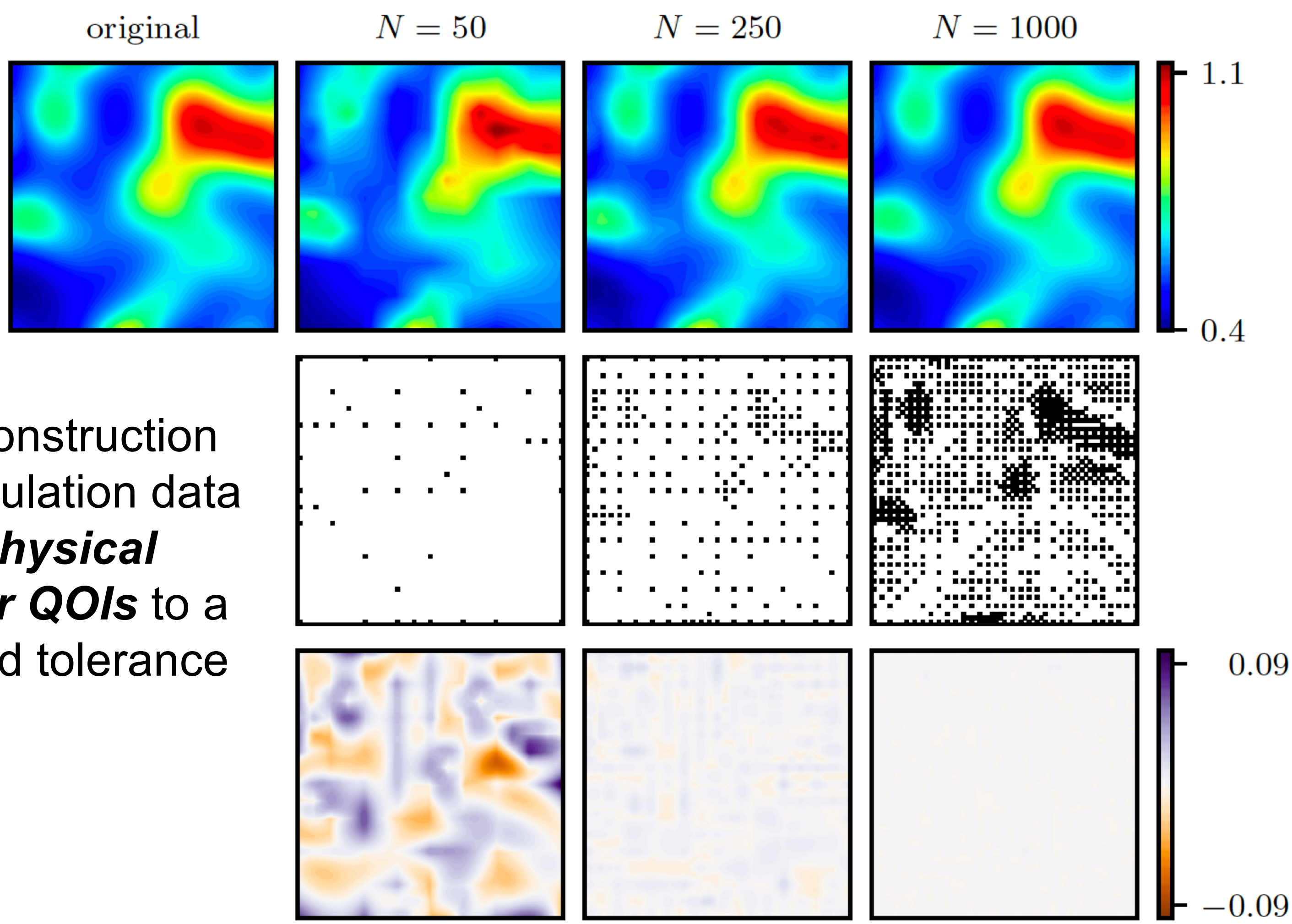
Research Details

- Approach allows transfer learning to new simulation data from the same application
- Online learning (using transfer learning) enables enhancing images from other application domains
- All metrics improved with machine learning: Normalized Mean Square Erorr (NMSE) reduced; Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR) increased



	NMSE	SSIM	PSNR
JPEG	0.038	0.971	37.245
CS w/ 400 iterations	0.045	0.973	34.534
EDSR	0.024	0.989	41.071
RDN	0.022	0.989	42.224

MGARD-Multigrid Adaptive Reduction of Data (Connections with RAPIDS and Un/Structured Grids)



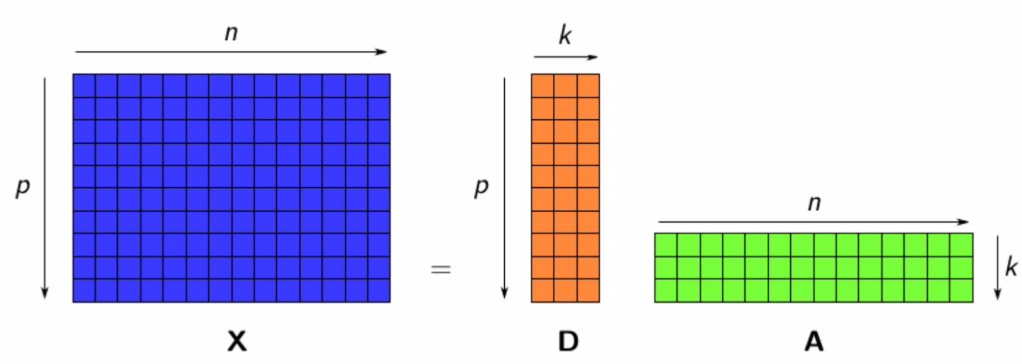
MGARD reconstruction of fusion simulation data preserving ***physical dynamics or QOIs*** to a pre-described tolerance level

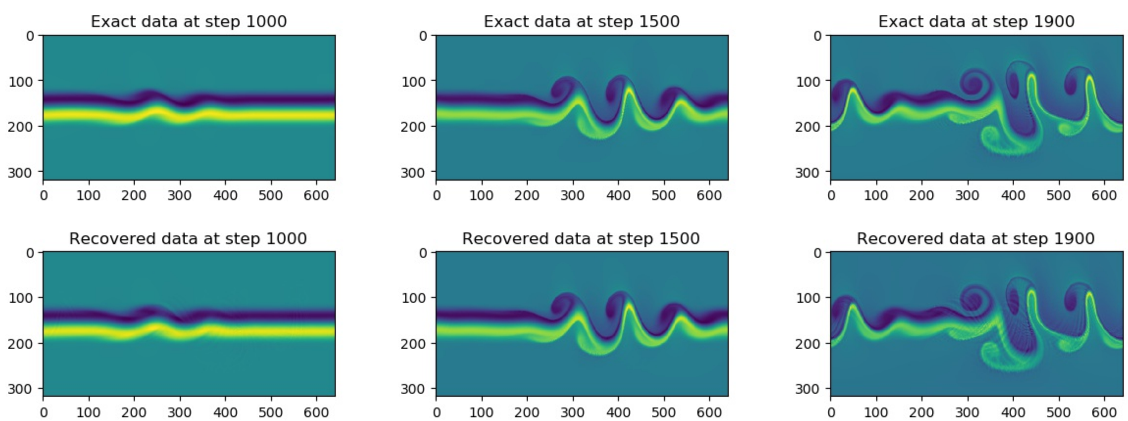
Compression and Reconstruction of Streaming Data (Connection to FASTMath Eigensolvers & Linear Solvers)

We develop a matrix factorization approach for data compression, reconstruction and interpretable decomposition:

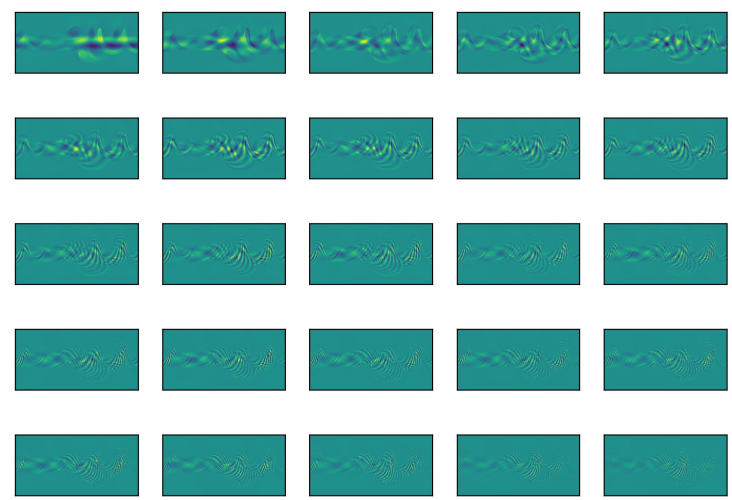
$$X \approx DA$$

- Data (signals, images) are stacked into $X \in \mathbb{R}^{p \times n}$.
- D: dictionary; A: sparse code.





Original and reconstructed data from online dictionary learning



Complete Dictionary